

From the Mechanics of the Mind to AGI: A Retrospective Review and Path Forward (2015–2025)

Dmitry A. Sklyarov

Abstract

This article presents a retrospective analysis of the “Mechanics of the Mind” concept (Sklyarov, 2015). We examine which of the proposed principles — including *proinformation*, *crossword memory*, the four brain algorithms, and emotions as efficiency signals — have already been partially or fully implemented in modern AI systems, and which remain unrealized. Despite significant progress in areas such as associative memory, world models, intrinsic motivation, and neuro-symbolic architectures, a holistic implementation of the concept is still lacking. We formulate architectural requirements for a system that could serve as the foundation for Artificial General Intelligence (AGI) based on the principles of the “Mechanics of the Mind.”

Note: This work is a direct continuation and critical review of the author’s earlier publication: D.A. Sklyarov, “The Mechanics of the Mind,” *Modeling of Artificial Intelligence*, vol. 8, no. 4, pp. 242–251, 2015.

Disclaimer: This article is a retrospective review of the author’s original Russian-language work (Sklyarov, 2015). The English translation was assisted by artificial intelligence tools. While every effort has been made to ensure accuracy, the author takes full responsibility for any inaccuracies or stylistic imperfections in the translation.

Keywords: Artificial General Intelligence, Proinformation, Crossword Memory, Brain Algorithms, Intrinsic Motivation, World Models, Neuro-Symbolic AI, Hierarchical Knowledge Formation, Configuration Simulation, Efficiency-Driven Emotions.

1 Introduction

2 Introduction

More than ten years have passed since the publication of the article “The Mechanics of the Mind” in 2015 — a period during which the field of artificial intelligence has undergone qualitative transformation. Transformer architectures, large-scale language models, neuro-symbolic hybrids, agents with internal world models, and systems capable of self-supervised learning in simulated environments have emerged. Nevertheless, despite impressive progress in narrow, specialized tasks, the challenge of building **Artificial General Intelligence (AGI)** — a mind capable of universal reasoning, self-development, and deep understanding of the world — remains unresolved.

In the 2015 work, the author proposed a conceptual model of the mind as the product of **algorithmic processing of proinformation** — an objective configuration of elements in the world, independent of any observer. Four core brain algorithms were formulated (perception, analogy, knowledge formation, configuration formation), the notion of **crossword memory** was introduced as the foundation of thought, and the role of **emotions as signals of processing efficiency** was demonstrated. These ideas were not offered as philosophical speculation but as a **practical engineering framework for constructing artificial mind**.

Today, in the era of rapid AI advancement, it has become possible to **evaluate the realizability of these principles**. Many of the proposed mechanisms — distributed knowledge storage, analogy-based reasoning, internal simulation, and motivation through uncertainty reduction — have already been partially embodied in modern architectures. However, they remain **fragmented**, lacking a unified ontological and architectural foundation.

The goal of this article is to conduct a **retrospective analysis of the “Mechanics of the Mind”** from the vantage point of 2025:

1. to identify which elements have already found reflection in AI practice,
2. to determine which principles remain unimplemented or underestimated,
3. to propose a pathway toward their integration into a unified architecture capable of serving as the basis for AGI.

Such an analysis not only confirms the foresight of the original concept but also underscores its **practical relevance** as a roadmap for the next generation of artificial intelligence.

3 Core Principles of the “Mechanics of the Mind”

The central ontological unit in the “Mechanics of the Mind” is **proinformation** — a term introduced to overcome the limitations of traditional definitions of information as “knowledge in context” (1). Unlike subjective interpretations, proinformation is defined as:

“a configuration of elements of the world — in static or dynamic states — represented fully or partially by means of configurations of other elements of the world” (1, p. 244).

This definition renders proinformation an **objective entity**, independent of observers. For example, genetic code is proinformation that defines an organism’s form and function, even if no one “reads” it. Similarly, the shape of a wave is proinformation — not the wave itself; a crystal lattice is proinformation — not the atoms.

Proinformation obeys physical laws: it cannot exist without a carrier and requires energy to change (in accordance with the conservation of energy principle). Its significance is characterized by **potential**, defined as:

$$\varphi_i = \frac{C_i}{K_i}, \tag{1}$$

where:

- φ_i — potential of the i -th unit of proinformation,
- C_i — number of connections (dependencies, objects described or controlled),
- K_i — number of copies of this proinformation.

When $\varphi_i = 1$, proinformation is in a state of “sufficiency”; $\varphi_i > 1$ indicates scarcity (e.g., a secret); $\varphi_i < 1$ indicates redundancy.

Based on proinformation, four fundamental brain algorithms are defined (1, pp. 246–248):

1. **Algorithm of Perception (AP)** — receives sensory input, decomposes it into proinformation fragments, and routes them for recognition.
2. **Algorithm of Analogy (AA)** — searches memory for similar configurations, guided by potential and data correspondence.

3. **Algorithm of Knowledge Formation (AKF)** — decomposes experience into “building blocks,” identifies unknowns, and constructs new structures (from shapes to scenes, sounds to narratives).
4. **Algorithm of Configuration Formation (ACF)** — in background mode (primarily during sleep), generates all valid combinations of known configurations to simulate possible futures and reinforce coherent links.

These algorithms operate not on isolated data but on a memory structured as a **3D crossword** (1, p. 248):

“Each cell (neuron) participates simultaneously in multiple intersecting ‘words’ (configurations). Meaning arises not in elements, but in their mutual intersection.”

This structure explains phenomena such as:

- **Attention** — a request for missing “letters” in the crossword,
- **Insight** — sudden completion of all cells,
- **Humor** — an emotional response to an unexpected yet logical solution that saves energy by sharply reducing uncertainty (1, pp. 248–249).

Importantly, the concept was explicitly intended as an **engineering foundation for AGI**. As stated in the conclusion:

“To create a self-developing artificial brain, it is unnecessary to replicate all features of the biological brain... it is necessary and sufficient to implement the principles of crossword memory, universal encoding, hierarchical knowledge formation, and bidirectional coupling with the world” (1, pp. 249–250).

These are not metaphors but **architectural prescriptions** — the focus of the following sections.

4 What Has Already Been Implemented in Modern AI (as of 2025)

Many principles from (1) now find partial embodiment in contemporary AI.

4.1 Crossword Memory and Associative Storage

The idea of memory as an intersecting, multi-contextual structure (1, p. 248) is reflected in:

- **Modern Hopfield Networks** (2), which store patterns as energy minima and retrieve via partial cues — analogous to “filling crossword cells”;
- **Graph and hypergraph neural networks**, where nodes participate in multiple relational contexts;
- **Embedding spaces of LLMs**, where semantic proximity reflects co-occurrence statistics — aligning with the notion of proinformation potential ($\varphi_i \propto \text{connectivity}$).

4.2 Algorithm of Analogy (AA)

The AA mechanism (1, p. 246) — search guided by potential and correspondence — underlies:

- **Attention in transformers**, which computes query-context relevance as structural analogy;
- **Retrieval-Augmented Generation (RAG)**, which explicitly retrieves analogous knowledge;
- **Few-shot learning**, where tasks are inferred by analogy to examples.

4.3 Algorithm of Knowledge Formation (AKF)

While not fully realized, components exist:

- **Neuro-symbolic systems** combine neural and symbolic processing to build compositional structures (e.g., “motion + shape = dance”);
- **Open-vocabulary perception** enables recognition of novel objects via description.

However, no system performs continuous, hierarchical knowledge construction “from scratch” as described in (1, p. 247).

4.4 Algorithm of Configuration Formation (ACF)

The ACF’s role in simulation and dreaming (1, p. 247) is echoed by:

- **World Models** (3), which learn internal dynamics for future prediction;

- **Generative replay** in reinforcement learning, which recombines experiences offline;
- **Dreaming agents** that hallucinate coherent scenarios for skill refinement.

4.5 Emotions as Efficiency Signals

The link between insight, energy savings, and reward (1, pp. 248–249) is mirrored in:

- **Intrinsic Curiosity Module (ICM)** (4), which rewards prediction error reduction;
- **Free Energy Principle** (5), which minimizes “surprise”;
- **Energy-efficient architectures**, which align with the brain’s drive to minimize resource use.

4.6 Proinformation as Objective Structure

Though the term is unused, the idea — world as observer-independent configuration — drives:

- **Causal AI** (e.g., (6; 7)), which seeks structural invariances;
- **Physics-informed neural networks (PINNs)**, which embed physical laws as constraints;
- **Geometric deep learning**, which treats data as points on manifolds with invariant properties.

Conclusion of Section 3: While individual mechanisms are implemented, no system integrates them into the unified cycle: $AP \rightarrow AA \rightarrow ACF \rightarrow AKF \rightarrow \text{action} \rightarrow \text{memory update}$.

5 What Remains Unimplemented

Despite progress, critical gaps persist.

5.1 No Unified Crossword Memory

Modern AI lacks a single, dynamic, intersecting memory substrate. LLM weights are static; RAG separates memory and model; no system supports real-time, multi-contextual reconfiguration of knowledge.

5.2 No Continuous Hierarchical Knowledge Construction

AI relies on fixed datasets and predefined feature spaces. It does not autonomously discover primitives or build knowledge hierarchically from streaming experience.

5.3 No Formal Integration of Proinformation

The ontology of observer-independent configurations is absent. There is no unified metric of significance (like φ_i) to guide attention, learning, or planning.

5.4 No Closed Cognitive Loop

Current systems decouple learning and inference, lack background simulation as a core process, and cannot actively query the world (e.g., “move camera to resolve ambiguity”).

5.5 No Integrated Motivational Contour

Internal rewards (curiosity, etc.) are not tied to a unified potential metric. Systems optimize for accuracy, not for *minimizing uncertainty at minimal energy cost*.

5.6 No Universal Encoding or Bidirectional World Coupling

Multimodal systems use separate encoders; most models are passive observers, not active agents that close the perception-action loop in real environments.

6 Proposal: A Path to AGI Based on the Mechanics of the Mind

We propose an architecture that unifies the principles of (1):

6.1 Unified Crossword Memory

A dynamic hypergraph where:

- each node (“letter”) participates in multiple configurations (“words”);
- edges are weighted by $\varphi_i = C_i/K_i$;
- structure grows hierarchically from primitives to abstractions.

6.2 Closed Loop of Four Algorithms

The system runs continuously:

1. AP: encodes multimodal input into proinformation fragments;
2. AA: retrieves analogous configurations using φ_i and correspondence;
3. ACF: fuses fragments into a unified percept;
4. AKF: extracts novel primitives and updates memory.

Action is generated; feedback updates memory — closing the loop.

6.3 Background Simulation (“Sleep”)

During idle periods, ACF generates valid combinations, tests coherence (physical, logical, causal), and stores robust hypotheses for planning and creativity.

6.4 Motivation via Proinformation Potential

Internal reward $r \propto -|\varphi_i - 1|$:

- $\varphi_i \rightarrow 1 \rightarrow$ “joy” $\rightarrow$ strengthen links;
- $\varphi_i \gg 1 \rightarrow$ curiosity $\rightarrow$ explore;
- $\varphi_i \ll 1 \rightarrow$ ignore.

6.5 Universal Encoding and World Coupling

All inputs (text, vision, sensors) map to a common proinformation graph. The agent acts in the world (speaks, moves, manipulates) and updates its model in real time.

6.6 Experimental Pathway

Start with a 2D agent:

1. Implement dynamic graph memory;
2. Train AP+AA to recognize objects via analogy;
3. Add AKF to discover new shapes;

4. Add ACF to simulate “what if I approach?”;
5. Use internal reward based on uncertainty reduction.

The agent *builds knowledge*, rather than learning from labels.

7 Conclusion

The “Mechanics of the Mind” (2015) was not merely prescient — it provided a **blueprint**. Modern AI has independently rediscovered many of its components, yet remains fragmented. The path to AGI lies not in scaling existing models, but in **synthesizing** proinformation, cross-word memory, the four algorithms, and efficiency-driven motivation into a single, self-developing architecture.

It should be noted that current attention mechanisms—such as those in transformers—function more like a microscope: they excel at high-resolution inspection of local context but lack the capacity for long-range, goal-directed exploration of knowledge space. To navigate across vast conceptual distances, a “telescope” is required: a dynamic focusing system akin to a tunable lens assembly that adjusts its focal depth based on the potential of proinformation and the structure of the crossword memory. Such an architecture would shift attention not merely by relevance, but by **semantic depth, structural coherence, and knowledge gaps**—enabling the system to “look ahead” rather than just “look closely.” The development of this next-generation attention mechanism will be the subject of a forthcoming publication.

Thus, the “Mechanics of the Mind” remains not merely a historical contribution to AI theory, but a **working roadmap toward AGI**.

As stated in the original work:

“It is necessary and sufficient to create a memory structure that satisfies the following principles...” (1, p. 250).

A decade later, the tools exist. The vision endures. The roadmap is clear.

References

- [1] Sklyarov, D. A. (2015). *The Mechanics of the Mind. Modeling of Artificial Intelligence*, 8(4), 242–251. Available at: https://archive.org/details/20251128_20251128_0859

- [2] Ramsauer, H., et al. (2021). Hopfield Networks is All You Need. *ICLR*.
- [3] Ha, D., & Schmidhuber, J. (2018). World Models. *arXiv:1803.10122*.
- [4] Pathak, D., et al. (2017). Curiosity-driven Exploration by Self-supervised Prediction. *CVPR Workshops*.
- [5] Friston, K. (2010). The Free Energy Principle. *Nature Reviews Neuroscience*, 11, 127–138.
- [6] Pearl, J. (2018). Theoretical Impediments to Machine Learning. *WSDM*.
- [7] Bengio, Y., et al. (2019). System 2 Deep Learning. *NeurIPS Workshop*.
- [8] Krotov, D., & Hopfield, J. J. (2016). Dense Associative Memory. *NeurIPS*.
- [9] Lake, B. M., et al. (2017). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences*, 40, e253.